

Topics in Markov chains: mixing and escape rate

Júlia Komjáthy and Yuval Peres

ABSTRACT. These are the notes for the minicourse on Markov chains delivered at the Saint Petersburg Summer School, June 2012. The main emphasis is on methods for estimating mixing times (for finite chains) and escape rates (for infinite chains). Lamplighter groups are key examples in both topics and the Varopolous-Carne long range estimate is useful in both settings.

1. Preliminaries

We start with preliminary notions necessary for the analysis of mixing and relaxation time of Markov chains. For much more on this topic see the books [AF02, LPW08].

1.1. Total variation distance and coupling. We start with the definition of total variation distance and coupling of two probability measures:

DEFINITION 1.1. Let \mathcal{S} be a state space, and μ and ν be two probability measures defined on \mathcal{S} . Then the **total variation distance** between μ and ν is defined as

$$\|\mu - \nu\|_{\text{TV}} = \max_{A \subset \mathcal{S}} |\mu(A) - \nu(A)|.$$

DEFINITION 1.2. A **coupling** of two probability measures μ and ν on \mathcal{S} is a pair of random variables (X, Y) having joint distribution q on $\mathcal{S} \times \mathcal{S}$ such that the marginal distributions are $\mathbb{P}[X = x] = \sum_{y \in \mathcal{S}} q(x, y) = \mu(x)$ and $\mathbb{P}[Y = y] = \sum_{x \in \mathcal{S}} q(x, y) = \nu(y)$ for every $x, y \in \mathcal{S}$.

1991 *Mathematics Subject Classification.* Primary 60J10, 16D05, 37A25.

Key words and phrases. Random walk, generalized lamplighter walk, wreath product, mixing time, relaxation time, Varopolous-Carne long range estimates.

J. Komjáthy was supported by the grant KTIA-OTKA # CNK 77778, funded by the Hungarian National Development Agency (NFÜ) from a source provided by KTIA.

Then, the followings give equivalent characterizations of the total variation distance $\|\mu - \nu\|_{\text{TV}}$:

$$(1.1) \quad \max_{A \subset \mathcal{S}} |\mu(A) - \nu(A)|$$

$$(1.2) \quad \frac{1}{2} \|\mu - \nu\|_1 = \frac{1}{2} \sum_{x \in \mathcal{S}} |\mu(x) - \nu(x)|$$

$$(1.3) \quad \sum_{x \in \mathcal{S}: \mu(x) > \nu(x)} (\mu(x) - \nu(x))$$

$$(1.4) \quad \inf \{ \mathbb{P}[X \neq Y] : (X, Y) \text{ is a coupling of } \mu \text{ and } \nu \}$$

PROOF. It is intuitively clear that the set $B := \{x : \mu(x) \geq \nu(x)\}$ or its complement maximizes the right hand side in Definition 1.1. To give a formal proof, take $A \subset \mathcal{S}$. From the definition of B it follows that

$$(1.5) \quad \mu(A) - \nu(A) \leq \mu(A \cap B) - \nu(A \cap B) \leq \mu(B) - \nu(B).$$

This proves that (1.1) \leq (1.3). But, if we take $A = B$, then the maximum is taken, i.e. (1.1) = (1.3). By the same reasoning, with $B^c := \mathcal{S} \setminus B$ we also have

$$(1.6) \quad \nu(A) - \mu(A) \leq \nu(A \cap B^c) - \mu(A \cap B^c) \leq \nu(B^c) - \mu(B^c).$$

Note that since $\mu(B^c) = 1 - \mu(B)$, $\nu(B^c) = 1 - \nu(B)$ the right hand side of (1.5) and (1.6) coincide, thus yielding

$$\max_{A \subset \mathcal{S}} |\mu(A) - \nu(A)| = \frac{1}{2} (\mu(B) - \nu(B) + \nu(B^c) - \mu(B^c)) = \frac{1}{2} \sum_{x \in \mathcal{S}} |\mu(x) - \nu(x)|,$$

proving (1.1) = (1.2).

To see that (1.1) \leq (1.4), we write

$$\begin{aligned} \mu(A) - \nu(A) &= \mathbb{P}[X \in A] - \mathbb{P}[Y \in A] \\ &\leq \mathbb{P}[X \in A, Y \notin A] \\ &\leq \mathbb{P}[X \neq Y]. \end{aligned}$$

For the other direction we construct a coupling for which the infimum is attained. Intuitively, what we do is pack as much mass into the diagonal $q(x, x)$ as we can, such that we still maintain the correct marginal measures. More formally, let us define

$$\begin{aligned} q(x, x) &:= \min\{\mu(x), \nu(x)\} \\ q(x, y) &:= 0 \text{ if } q(x, x) = \mu(x) \text{ or } q(y, y) = \nu(y) \\ q(x, y) &= \frac{(\mu(x) - \nu(x))(\nu(y) - \mu(y))}{1 - \sum_z q(z, z)} \text{ if } q(x, x) = \nu(x) \text{ and } q(y, y) = \mu(y). \end{aligned}$$

Intuitively, we put the maximal possible weight in the diagonal of q , (which is $\min\{\mu(x), \nu(x)\}$) and then we put zeros in the corresponding column or row, depending on the minimum being $\mu(x)$ or $\nu(x)$. Finally, we fill the rest out with conditionally independent choice, i.e. on $B \times B^c$ we distribute $(\mu(x) - \nu(x)) \cdot (\nu(y) - \mu(y)) > 0$ with the normalizing factor $1 - \sum_z q(z, z)$. Mind that this is not the only way of doing the coupling. To check that the marginals are correct is left to the reader.

With this particular coupling, (1.4) becomes

$$\begin{aligned}
 (1.4) \leq \mathbb{P}(X \neq Y) &= 1 - \sum_x q(x, x) = 1 - \sum_x \min\{\mu(x), \nu(x)\} \\
 &= \sum_x \mu(x) - \left(\sum_{x: \mu(x) > \nu(x)} \nu(x) + \sum_{x: \mu(x) \leq \nu(x)} \mu(x) \right) \\
 &= \sum_{x: \mu(x) > \nu(x)} [\mu(x) - \nu(x)] = (1.3).
 \end{aligned}$$

With this we have (1.4) \leq (1.3) $=$ (1.1), finishing the proof. \square

1.2. Mixing in total variation distance. Let X_t be a Markov chain on state space \mathcal{S} with transition matrix P , and stationary measure π on \mathcal{S} . That is, $\pi P = \pi$. If P is irreducible and aperiodic, then the measure $\mu_t(y) = P^t(x, y)$ is converging to the stationary measure exponentially fast, i.e. there exists an $\alpha \in (0, 1)$ such that

$$\|P^t(x, \cdot) - \pi(\cdot)\|_{TV} \leq C\alpha^t.$$

These asymptotics hold for a single chain as the time t tends to infinity. However, we are rather interested in the finite time behavior of a sequence of Markov chains, i.e. how long one has to run the Markov chain as a function of $|\mathcal{S}|$, to get ε -close to stationary measure, for fixed ε .

Thus, let us define

$$(1.7) \quad d_x(t) := \|P^t(x, \cdot) - \pi(\cdot)\|_{TV}; \quad d(t) := \max_{x \in \mathcal{S}} d_x(t).$$

Then, the ε -mixing time of a Markov Chain on a graph G is defined as

$$(1.8) \quad t_{\text{mix}}(G, \varepsilon) := \min \{t \geq 0 : d(t) \leq \varepsilon\}.$$

Throughout, we set $t_{\text{mix}}(G) := t_{\text{mix}}(G, \frac{1}{4})$. The characterisation (1.4) suggests that sometimes it is more convenient to work with chains started from two different initial states, so let us define

$$\bar{d}(t) := \max_{x, y \in \mathcal{S}} \|P^t(x, \cdot) - P^t(y, \cdot)\|_{TV}.$$

Then, we have the following comparison:

LEMMA 1.3. *With the above definitions,*

$$(1.9) \quad d(t) \leq \bar{d}(t) \leq 2d(t)$$

Further, the function $\bar{d}(t)$ is submultiplicative, i.e.

$$(1.10) \quad \bar{d}(t+s) \leq \bar{d}(t)\bar{d}(s),$$

and combining yields

$$(1.11) \quad d(kt) \leq 2^k d(t)^k$$

PROOF. We only prove (1.9) here. The proof of (1.10) is the proof of Lemma 4.12 in [LPW08], and (1.11) is an easy combination of the first two statements of the lemma. To prove the second inequality in (1.9), we use the triangle inequality

$$\begin{aligned}
 \bar{d}(t) &= \|P^t(x, \cdot) - P^t(y, \cdot)\|_{TV} \leq \|P^t(x, \cdot) - \pi(\cdot)\|_{TV} + \|\pi(\cdot) - P^t(y, \cdot)\|_{TV} \\
 &\leq 2d(t),
 \end{aligned}$$

and for the first inequality we can use that $\pi P^t = \pi$ to get

$$d_x(t) = \max_{A \in \mathcal{S}} |P^t(x, A) - \pi(A)| = \max_{A \in \mathcal{S}} \left| \sum_{y \in \mathcal{S}} \pi(y) P^t(x, A) - P^t(y, A) \right|.$$

Now, by the triangle inequality the right hand side is at most

$$\begin{aligned} \max_{A \in \mathcal{S}} \sum_{y \in \mathcal{S}} \pi(y) |P^t(x, A) - P^t(y, A)| &\leq \sum_{y \in \mathcal{S}} \pi(y) \max_{A \in \mathcal{S}} |P^t(x, A) - P^t(y, A)| \\ &= \bar{d}(t). \end{aligned}$$

□

The definition $\bar{d}(t)$ is extremely useful, since it allows us to relate the mixing time of the chain to the tail behavior of the so called **coupling time**: Given a coupling (X_t, Y_t) of $P^t(x, \cdot)$ and $P^t(y, \cdot)$, let us define

$$\tau_{\text{couple}} := \min\{t : X_t = Y_t\}.$$

Then we have

$$(1.12) \quad d(t) \leq \bar{d}(t) \leq \max_{x, y} \mathbb{P}[X_t \neq Y_t] = \max_{x, y} \mathbb{P}[\tau_{\text{couple}} > t].$$

With all these prerequisites in our hands, we can state and prove our first theorem:

THEOREM 1.4. *The mixing time of C_n , the cycle on n vertices is bounded from above by*

$$t_{\text{mix}}(C_n) \leq n^2.$$

PROOF. We will construct a coupling of the measures $P^t(x, \cdot)$ and $P^t(y, \cdot)$ and use (1.12) to estimate $d(t)$. Note that $P^t(x, \cdot)$ and $P^t(y, \cdot)$ are the transition measures of two lazy random walks, say X_t and Y_t , with $X_0 = x$ and $Y_0 = y$. Thus, we construct a coupling of (X_t, Y_t) as follows: we couple the increments of the walks, as long as $X_t \neq Y_t$ holds:

$$\begin{aligned} \mathbb{P}(X_t - X_{t-1} = 0, Y_t - Y_{t-1} = +1) &= \frac{1}{4}; \quad \mathbb{P}(X_t - X_{t-1} = 0, Y_t - Y_{t-1} = -1) = \frac{1}{4}; \\ \mathbb{P}(X_t - X_{t-1} = +1, Y_t - Y_{t-1} = 0) &= \frac{1}{4}; \quad \mathbb{P}(X_t - X_{t-1} = -1, Y_t - Y_{t-1} = 0) = \frac{1}{4}. \end{aligned}$$

If the two walks meet then they stay together from that point on. It is easy to check that the marginals of the two walks are correct. The advantage of this coupling is that before collision the two walks never move at the same time. I.e., the clockwise distance $D_t = X_t - Y_t$ changes at each step by $+1$ or -1 . This means that D_t is doing a simple (non-lazy) symmetric random walk on $\{0, 1, \dots, n\}$ with $D_0 := k \in \{1, \dots, n-1\}$, and we are waiting until it hits 0 or n . This is exactly the well known *Gambler's ruin* problem. The coupling time is then $\tau_{0,n}$, the hitting time of the set $\{0, n\}$. We can use the martingale D_t and use optional stopping to calculate its expected value:

$$k = \mathbb{E}_k[D_0] = \mathbb{E}_k[D_{\tau_{0,n}}] = \mathbb{P}_k[D_{\tau_{0,n}} = n]n,$$

from which $\mathbb{P}_k[D_{\tau_{0,n}} = n] = k/n$. Then, $D_t^2 - t$ is also a martingale, (to check is left for the reader as an exercise) and using the previous calculation and optional stopping gives

$$\mathbb{E}_k[\tau_{0,n}] = k(n-k) \leq n^2/4.$$

Combining Lemma 1.3, the characterisation (1.4) of the total variation distance and the previous calculations with a Markov's inequality, we arrive at the following sequence of inequalities:

$$d(t) \leq \bar{d}(t) \leq \max_{x,y} \mathbb{P}[X_t \neq Y_t] = \max_k \mathbb{P}_k[D_t > t] \leq \frac{\mathbb{E}_k[D_t]}{t} \leq \frac{n^2}{4t}.$$

Now let us set $t = n^2$, then we get $d(n^2) \leq 1/4$, implying $t_{\text{mix}}(C_n) \leq n^2$. \square

A similar coupling can be used to give an upper bound on the mixing time on the d -dimensional tori:

THEOREM 1.5. *The total variation mixing time on \mathbf{Z}_n^d , the d -dimensional torus is bounded from above by*

$$(1.13) \quad t_{\text{mix}}(\mathbf{Z}_n^d) \leq 3d \log d \, n^2.$$

PROOF. We couple the two walks $\underline{X}_t = (X_t^1, X_t^2, \dots, X_t^d)$ and $\underline{Y}_t = (Y_t^1, Y_t^2, \dots, Y_t^d)$ coordinate-wise with the same coupling as in the proof of Theorem 1.4. More precisely, at each step we first pick a uniform number $U_t \in \{1, 2, \dots, d\}$ independently of everything else, and then, we check if the corresponding coordinates $X_t^{U_t}, Y_t^{U_t}$ coincide or not. If so, we move both walks with the same increment: 0, +1 or -1 with probabilities 1/2, 1/4, 1/4 each. If $X_t^{U_t} \neq Y_t^{U_t}$, then we apply the coupling described in the proof of Theorem 1.4 for the U_t th coordinate. Let D_t^i denote the clockwise difference between X_t^i and Y_t^i , and τ_i denote the first time when D_t^i hits $\{0, n\}$. Since each coordinate i has a Geometric($1/d$) waiting time for its next move, the marginal distribution of each τ_i can be written as

$$\tau_i = \sum_{j=1}^{\tau_{0,n}^{(i)}} Z_j$$

with $Z_j \sim \text{Geo}(1/d)$, and $\tau_{0,n}^{(i)} \sim \tau_{0,n}$ as in the proof of Theorem 1.4. This gives that $\mathbb{E}[\tau_i] \leq \frac{dn^2}{4}$. Note that this bound holds for every starting point x_i, y_i . So we can run the chain in blocks of $dn^2/2$ and then in each block we hit the set $\{0, n\}$ with probability at least 1/2 by Markov's inequality. Hence the hitting of the set $\{0, n\}$ is stochastically dominated by a random variable of the form $\frac{1}{2}dn^2\text{Geo}(1/2)$. This yields the bound

$$\mathbb{P}[\tau_i > t] \leq 2 \left(\frac{1}{2} \right)^{\frac{2t}{dn^2}},$$

where the factor 2 comes from ignoring the integer part of $\frac{2t}{dn^2}$. Set $t = 3d \log d \cdot n^2$, then, for all $d \geq 2$:

$$\mathbb{P}[\underline{X}_t \neq \underline{Y}_t] = \mathbb{P}[\exists i : \tau_i > t] \leq d \cdot \mathbb{P}[\tau_i > t] \leq 2d \left(\frac{1}{2} \right)^{\frac{2t}{dn^2}} = 2d^{1-6 \log 2} \leq \frac{1}{4}.$$

Hence we have $t_{\text{mix}}(\mathbf{Z}_n^d) \leq 3d \log d \, n^2$, finishing the proof. \square

1.3. Strong stationary times. In many cases the following random times give a useful bound on mixing times:

DEFINITION 1.6. A randomized stopping time τ is called a strong stationary time for the Markov chain X_t on G if

$$(1.14) \quad \mathbb{P}_x[X_\tau = y, \tau = t] = \pi(y) \mathbb{P}_x[\tau = t],$$

that is, the position of the walk when it stops at τ is distributed according to π and independent of the value of τ .

The adjective randomized means that the stopping time can depend on some extra randomness, not just purely the trajectories of the Markov chain, for a precise definition see [LPW08, Section 6.2.2].

DEFINITION 1.7. A state $h(x) \in V(G)$ is called a *halting state* for a stopping time τ and initial state x if $\{X_t = h(x)\}$ implies $\{\tau \leq t\}$.

Strong stationary times are useful since they are closely related to an other notion of distance from the stationary measure. We define

DEFINITION 1.8. The *separation distance* $s(t)$ is defined as

$$(1.15) \quad s(t) := \max_{x \in \mathcal{S}} s_x(t) \text{ with } s_x(t) := \max_{y \in \mathcal{S}} \left(1 - \frac{P^t(x, y)}{\pi(y)} \right).$$

We mention that the separation distance is not a metric.

The relation between the separation distance and any strong stationary time τ is the following inequality from [AF02] or [LPW08, Lemma 6.11]:

$$(1.16) \quad \forall x \in \mathcal{S} : s_x(t) \leq \mathbb{P}_x(\tau > t).$$

The proof is just two lines, so we include it here for the reader's convenience: for any y we have

$$(1.17) \quad 1 - \frac{P^t(x, y)}{\pi(y)} \leq 1 - \frac{\mathbb{P}_x[X_t = y, \tau \leq t]}{\pi(y)}$$

Now (1.14) implies that the last expression equals

$$1 - \frac{\pi(y)\mathbb{P}_x[\tau \leq t]}{\pi(y)} = \mathbb{P}_x[\tau > t].$$

Later we will need a slightly stronger result than (1.16), namely from (1.17) it follows that if τ has a halting state $h(x)$ for x , then putting $y = h(x)$ yields that equality holds in (1.16). Unfortunately, the statement can not be reversed: the state $h(x, t)$ maximizing the separation distance at time t can also depend on t and thus the existence of a halting state is not necessarily needed to get equality in (1.16).

On the other hand, one can always construct τ such that (1.16) holds with equality for every $x \in \mathcal{S}$. This τ does not necessarily obeys halting states. This is one of the main ingredients to our proofs in Section 2, so we cite it as a Theorem (with adjusted notation).

THEOREM 1.9. [Aldous, Diaconis]/[AD86, Proposition 3.2] Let $(X_t, t \geq 0)$ be an irreducible aperiodic Markov chain on a finite state space \mathcal{S} with initial state x and stationary distribution π , and let $s_x(t)$ be the separation distance defined as in (1.15). Then

- (1) if τ is a strong stationary time for X_t , then $s_x(t) \leq \mathbb{P}_x(\tau > t)$ for all $t \geq 0$.
- (2) Conversely, there exists a strong stationary time τ such that $s_x(t) = \mathbb{P}_x(\tau > t)$ holds with equality.

Combining these, we will call a strong stationary time τ *separation optimal* if it achieves equality in (1.16). Mind that every stopping time possessing halting states is separation optimal, but not the other way round.

The next lemma relates the total and the separation distance:

LEMMA 1.10. *For any reversible Markov chain and any state $x \in \mathcal{S}$, the separation distance from initial vertex x satisfies:*

$$(1.18) \quad d_x(t) \leq s_x(t)$$

$$(1.19) \quad s_x(2t) \leq 4d(t)$$

PROOF. For a short proof of (1.18) see [AF02] or [LPW08, Lemma 6.13], and combine [LPW08, Lemma 19.3] with a triangle inequality to conclude (1.19). Here we write the proofs for the reader's convenience. We have

$$\begin{aligned} d_x(t) &= \sum_{\substack{y \in \mathcal{S} \\ P^t(x,y) < \pi(y)}} [\pi(y) - P^t(x,y)] = \sum_{\substack{y \in \mathcal{S} \\ P^t(x,y) < \pi(y)}} \pi(y) \left[1 - \frac{P^t(x,y)}{\pi(y)} \right] \\ &\leq \max_y \left[1 - \frac{P^t(x,y)}{\pi(y)} \right] = s_x(t). \end{aligned}$$

To see (1.19), we mind that reversibility means that $P^t(z,y)/\pi(y) = P^t(y,z)/\pi(z)$. Hence we have

$$\frac{P^{2t}(x,y)}{\pi(y)} = \sum_{z \in \mathcal{S}} \frac{P^t(x,z)P^t(z,y)}{\pi(y)} = \sum_{z \in \mathcal{S}} \frac{P^t(x,z)P^t(y,z)}{\pi(z)} \cdot \sum_{z \in \mathcal{S}} \pi(z)$$

Applying Cauchy-Schwarz to the right hand side implies

$$\frac{P^{2t}(x,y)}{\pi(y)} \geq \left(\sum_{z \in \mathcal{S}} \sqrt{P^t(x,z)P^t(y,z)} \right)^2 \geq \left(\sum_{z \in \mathcal{S}} P^t(x,z) \wedge P^t(y,z) \right)^2.$$

Recall (1.4), i.e.

$$\sum_z \mu(z) \wedge \nu(z) = 1 - \|\mu - \nu\|_{\text{TV}}.$$

Combining this with the previous calculation results in

$$1 - \frac{P^{2t}(x,y)}{\pi(y)} \leq 1 - (1 - \|P^t(x, \cdot), P^t(y, \cdot)\|_{\text{TV}})^2.$$

Using the triangle inequality $\|P^t(x, \cdot) - P^t(y, \cdot)\|_{\text{TV}} \leq 2d(t)$ and expanding the terms yields (1.19). \square

We demonstrate the use of strong stationary times by analysing the *separation time* of the d -dimensional hypercube: the separation time is defined similarly as the mixing time in (1.8) by replacing $d(t)$ by $s(t)$.

THEOREM 1.11. *For the lazy random walk on the hypercube $H_d = \{0, 1\}^d$,*

$$t_{\text{sep}}(H_d, \varepsilon) \leq d \log d + \log(1/\varepsilon)d.$$

PROOF. We construct the following strong stationary time for the lazy random walk on the hypercube: independently in each step, we pick a uniform coordinate $U_t \in \{0, 1, \dots, d\}$, and then independently of the current values and everything else, we set $X_t^{U_t} = 1$ with probability $1/2$ and $X_t^{U_t} = 0$ with probability $1/2$. By doing so, the probability that the chain stays put is exactly $1/2$, and with probability $1/2$

it moves to a position chosen uniformly among all neighboring vertices, i.e., we get exactly the transition probabilities for a lazy random walk on the hypercube.

Define τ_{refresh} as the first time that all coordinates have been chosen. Then, at τ_{refresh} , each coordinate $i \in \{1, \dots, d\}$ has been selected already at least once, thus, its position is 0 or 1 with probability $1/2$ each, independently of how long we had to wait for τ_{refresh} to happen. Also, if the original state was $\underline{x} = (x_1, x_2, \dots, x_d)$, then to reach $h(\underline{x}) = (1 - x_1, 1 - x_2, \dots, 1 - x_d)$, we have to refresh each coordinate at least once, i.e., $h(\underline{x})$ is a halting state for τ_{refresh} . This shows that τ_{refresh} is a *separation-optimal strong stationary time* for the lazy RW on the hypercube.

Note that the distribution of τ_{refresh} is the same as that of the *coupon collector* problem:

$$s_x(t) = \mathbb{P}_x[\tau_{\text{refresh}} > t] = \mathbb{P}[\exists i \in \{1, \dots, d\} : \forall s \leq t \ U_s \neq i] \leq d \left(1 - \frac{1}{d}\right)^t.$$

By putting $t = d \log d - \log(\varepsilon)d$, the right hand side of the previous display is less than $e^{\log(\varepsilon)} = \varepsilon$, finishing the proof. \square

REMARK 1.12. It is known (see [LPW08, Example 12.17] that the *total variation mixing time* of the hypercube is at $\frac{1}{2}d \log d$, hence we have a factor 2 between the separation and tv-mixing time on H_d . Comparing it to the estimate in (1.19), this shows that the factor 2 there can be sharp.

The following lemma will be used later to determine the spectral gap of the lamplighter chain: ([LPW08, Corollary 12.6])

LEMMA 1.13. *For a reversible, irreducible and aperiodic Markov chain,*

$$(1.20) \quad \begin{aligned} d_x(t) &\leq s_x(t) \leq \frac{\lambda_*^t}{\pi_{\min}}, \\ |\lambda_2|^t &\leq 2d(t) \end{aligned}$$

with $\pi_{\min} = \min_{y \in \mathcal{S}} \pi(y)$ and $\lambda_* = \max\{|\lambda| : \lambda \text{ eigenvalue of } P, \lambda \neq 1\}$. As a consequence we have

$$\lim_{t \rightarrow \infty} d(t)^{1/t} = \lambda_*.$$

PROOF. Follows from [LPW08, Equation (12.11), (12.13)]. We note that Lemma 1.10 implies that the assertion of Lemma 1.13 stays valid if we replace $d(t)^{1/t}$ by the separation distance $s(t)^{1/t}$. \square

2. Mixing times of lamplighter graphs

In this section, we will use the preliminaries from the previous sections to determine the mixing and relaxation time of the random walk on lamplighter graphs. The intuitive representation of the walk is the following: a lamplighter moves according to a simple random walk on the vertices of a *base graph* G . Further, there is an identical lamp attached to each vertex $v \in G$, and each of the lamps is either on or off. We denote the state of the lamp at vertex $v \in G$ by f_v . Then, as the lamplighter walks along the base graph, he switches on or off lamps on its path randomly. More precisely, we are analyzing the following dynamics below: one move of the lamplighter walk corresponds to three elementary steps: he randomizes the lamp on its current position, then he moves according to a lazy simple random walk on the base graph, then he randomizes the lamp at its arrival position.

Suppose that G is a finite connected graph with vertices $V(G)$ and edges $E(G)$. We refer to G as the base graph. Let $\mathcal{X}(G) = \{\underline{f}: V(G) \rightarrow \{0, 1\}\}$ be the set of markings of $V(G)$ by elements of $\{0, 1\}$. The wreath product $\mathbf{Z}_2 \wr G$ is the graph whose vertices are pairs (\underline{f}, x) where $\underline{f} = (f_v)_{v \in V(G)} \in \mathcal{X}(G)$ and $x \in V(G)$. There is an edge between (\underline{f}, x) and (\underline{g}, y) if and only if $(x, y) \in E(G)$, and $f_z = g_z$ for all $z \notin \{x, y\}$. Suppose that P is the transition matrix for lazy random walk on G . The lamplighter walk X^\diamond is the Markov chain on $\mathbf{Z}_2 \wr G$ which moves from a configuration (\underline{f}, x) by

- (1) picking y adjacent to x in G according to P , then
- (2) updating each of the values of f_x and f_y independently to a uniform random value in $\{0, 1\}$.

The state of lamps f_z at all other vertices $z \in G$ remain fixed. It is easy to see that with stationary distribution π_G for the random walk on G , the unique stationary distribution of X^\diamond is the product measure

$$\pi^\diamond((\underline{f}, x)) = \pi_G(x) \cdot 2^{-|G|},$$

and X^\diamond is itself reversible. In this notes, we will be concerned with the special case that P is the transition matrix for the *lazy random walk* on G . In particular, P is given by

$$(2.1) \quad P(x, y) := \begin{cases} \frac{1}{2} & \text{if } x = y, \\ \frac{1}{2d(x)} & \text{if } \{x, y\} \in E(G), \end{cases}$$

for $x, y \in V(G)$ and where $d(x)$ is the degree of x . This assumption guarantees that we avoid issues of periodicity.

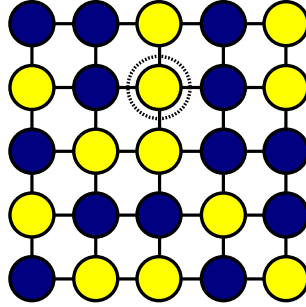


FIGURE 1. A typical state of the lamplighter walk on the 2-dim torus on 5 vertices. Lamps that are ‘on’ (resp. ‘off’) are drawn in yellow (resp. blue) and the position of the lamplighter is marked by the dashed circle.

We will study below the *total variation (TV)* mixing time and the *relaxation time* of these walks. The relaxation time is a more algebraic point of view of mixing, that looks at the spectral behavior of the transition matrix P . Namely, since P is a stochastic matrix, 1 is the main eigenvalue and all the other eigenvalues of P lie in the complex unit disk. If further the chain is reversible, then the eigenvalues are real and it makes sense to define the relaxation time of the chain by

$$t_{\text{rel}}(G) := \frac{1}{1 - \lambda_2},$$

where λ_2 is the second largest eigenvalue of the chain.

In general it is known that for a reversible Markov chain the asymptotic behavior of the relaxation time, the TV and a third notion, the *uniform* mixing time, which is mixing in ℓ_∞ norm, can significantly differ, i.e. in terms of the size of the graph G they can have different asymptotics. More precisely, we have

$$t_{\text{rel}}(G) \leq t_{\text{mix}}^{TV}(G, 1/4) \leq t_{\text{mix}}^u(G, 1/4),$$

see [AF02] or [LPW08]. The lamplighter walk described above is an example where these three quantities have different order of magnitude in terms of $|G|$.

Throughout, we use the superscript \diamond to specify that a quantity belongs to the lamplighter walk, that is, the underlying graph is $\mathbf{Z}_2 \wr G$. In order to state our general theorems, we first need to review some basic terminology from the theory of Markov chains. Let P be the transition kernel for a lazy random walk on a finite, connected graph G with stationary distribution π .

The *maximal hitting time* of P is

$$(2.2) \quad t_{\text{hit}}(G) := \max_{x, y \in V(G)} \mathbf{E}_x[\tau_y],$$

where τ_y denotes the first time t that $X(t) = y$ and \mathbf{E}_x stands for the expectation under the law in which $X(0) = x$. The random cover time τ_{cov} is the first time when all vertices have been visited by the walker X , and the cover time $t_{\text{cov}}(G)$ is

$$(2.3) \quad t_{\text{cov}}(G) := \max_{x \in V(G)} \mathbf{E}_x[\tau_{\text{cov}}].$$

Then we have the following two theorems (from [PR04]):

THEOREM 2.1. *Let us assume that G is a regular, connected graph. Then there exist universal constants $0 < c_1 \leq C_1 < \infty$ such that the relaxation time of the lamplighter walk on $\mathbf{Z}_2 \wr G$ satisfies*

$$(2.4) \quad c_1 t_{\text{hit}}(G) \leq t_{\text{rel}}(\mathbf{Z}_2 \wr G) \leq C_1 t_{\text{hit}}(G),$$

THEOREM 2.2. *Let G be a regular connected graph. Then there exist universal constants $0 < c_2 \leq C_2 < \infty$ such that the mixing time of the lamplighter walk on $\mathbf{Z}_2 \wr G$ satisfies*

$$(2.5) \quad c_2 t_{\text{cov}}(G) \leq t_{\text{mix}}(\mathbf{Z}_2 \wr G) \leq C_2 t_{\text{cov}}(G).$$

2.1. Proofs. Here we modify the proof that can be found in [KP12] for more general lamp graphs to the setting where the lamp graph is \mathbf{Z}_2 . We start by constructing an ‘almost’ stationary time τ^\diamond for the lamplighter walk. More specifically, the first refreshment of a lamp at site v is a strong stationary time on the copy at v of the two-state Markov chain on $\{0, 1\}$, and we stop the chain when all lamps reach their individual stopping time, i.e. exactly when we cover all vertices. At τ_{cov} , the lamps are already stationary, but the position of the walker not necessarily.

It is easy to see that the state of the lamps are already stationary when τ_{cov} has happened, that is, for any starting state (\underline{f}_0, x_0)

$$(2.6) \quad \mathbb{P}_{(\underline{f}_0, x_0)}[X_t^\diamond = (\underline{f}, x), \tau_{\text{cov}} = t] = 2^{-|G|} \cdot \mathbb{P}_{(\underline{f}_0, x_0)}[X_t = x, \tau_{\text{cov}} = t].$$

Further, if a lamp is in state x , then $1 - x$ is a halting state for the two state Markov chain. From this it is not hard to see that the vectors $((1 - f_0(v))_{v \in G}, y)$ are halting state vectors for τ_{cov} and initial state (\underline{f}_0, x_0) for every $x_0, y \in G$.

LEMMA 2.3. *For the separation distance on the lamplighter chain $\mathbf{Z}_2 \wr G$ the following lower bound holds:*

$$s_{(\underline{f}_0, x_0)}^\diamond(t) \geq \mathbb{P}_{(\underline{f}_0, x_0)}[\tau_{\text{cov}} > t].$$

PROOF. Observe that reaching the halting state vector $((1 - f_0(v))_{v \in G}, x)$ implies the event $\tau_{\text{cov}} \leq t$ so we have

$$(2.7) \quad \frac{\mathbb{P}_{(\underline{f}_0, x_0)}[X_t^\diamond = ((1 - f_0(v))_{v \in G}, x)]}{\pi_G(x)2^{-|G|}} = \frac{\mathbb{P}_{(\underline{f}_0, x_0)}[X_t^\diamond = ((1 - f_0(v))_{v \in G}, x), \tau_{\text{cov}} \leq t]}{\pi_G(x)2^{-|G|}}$$

Now pick a vertex $x = x_{x_0, t} \in G$ which minimizes $\mathbb{P}[X_t = x_{x_0, t} | \tau_{\text{cov}} \leq t] / \pi_G(x_{x_0, t})$. This quotient is less than 1 since both the numerator and the denominator are probability distributions on G . Then, using this and (2.6), 1 minus the right hand side of (2.7) equals

$$1 - \frac{\mathbb{P}_{(\underline{f}_0, x_0)}[X_t = x_{x_0, t} | \tau^\diamond \leq t] \mathbb{P}_{(\underline{f}_0, x_0)}[\tau^\diamond \leq t]}{\pi_G(x_{x_0, t})} \geq 1 - \mathbb{P}_{(\underline{f}_0, x_0)}[\tau^\diamond \leq t].$$

The separation distance is larger than the left hand side of (2.7) by definition, and the proof of the claim follows. \square

With this lemma in hand, we can already prove the lower bound in Theorem 2.2.

PROOF OF THE LOWER BOUND FOR MIXING TIME OF $\mathbf{Z}_2 \wr G$. Let us set $t := 6t_{\text{mix}}(\mathbf{Z}_2 \wr G)$. Then Lemma 2.3 and Lemma 1.10 yields us the following sequence of inequalities:

$$\mathbb{P}_{(\underline{f}_0, x_0)}[\tau_{\text{cov}} > 6t_{\text{mix}}(\mathbf{Z}_2 \wr G)] \leq s_{(\underline{f}_0, x_0)}^\diamond(6t_{\text{mix}}(\mathbf{Z}_2 \wr G)) \leq 4d^\diamond(3t_{\text{mix}}(\mathbf{Z}_2 \wr G)) \leq \frac{1}{2},$$

where in the last inequality we used the sub-multiplicativity property (1.11). Note that this estimate is independent of the starting state. Comparing the left and right hand sides, we conclude that we can run the chain in blocks of $6t_{\text{mix}}(\mathbf{Z}_2 \wr G)$, and in each block the graph G is covered with probability at least 1/2. Thus, τ_{cov} can be stochastically dominated by $6t_{\text{mix}}(\mathbf{Z}_2 \wr G)\text{Geo}(1/2)$. Taking expected value yields

$$t_{\text{cov}}(G) \leq 12t_{\text{mix}}(\mathbf{Z}_2 \wr G),$$

finishing the lower bound with $c_2 = 1/12$. \square

PROOF OF THE UPPER BOUND FOR MIXING TIME OF $\mathbf{Z}_2 \wr G$. The proof of the upper bound in Theorem 2.2 is very similar, we just need to make the position of the lamplighter also stationary. We can achieve this by waiting an extra strong stationary time τ_G after $\tau^\diamond \equiv \tau_{\text{cov}}$ has happened. The existence of a *separation optimal* strong stationary time on G is ensured by Theorem 1.9.

More precisely, we have

LEMMA 2.4. *Let $\tau_G(x)$ be a separation-optimal strong stationary time for G starting from $x \in G$ and define τ_2^\diamond by*

$$(2.8) \quad \tau^\diamond := \tau_{\text{cov}} + \tau_G(X_{\tau_{\text{cov}}}),$$

where the chain is re-started at τ_{cov} from $(\underline{E}_{\tau^\diamond}, X_{\tau^\diamond})$, run independently of the past and τ_G is measured in this walk. Then, τ^\diamond is a strong stationary time for $\mathbf{Z}_2 \wr G$.

The proof of this lemma is omitted here since it is not difficult but quite long, see [KP12].

With this lemma in hand, we can apply (1.16) – the relation between separation distance and strong stationary times – to get

$$(2.9) \quad d_{(\underline{f}_0, x_0)}^\diamond(t) \leq s_{(\underline{f}_0, x_0)}^\diamond(t) \leq \mathbb{P}_{(\underline{f}_0, x_0)}[\tau_{\text{cov}} + \tau_G(X_{\tau_{\text{cov}}}) > t].$$

Now set $t = 8t_{\text{cov}}(G) + 10t_{\text{mix}}(G)$. Then by a union bound the right hand side in (2.9) is at most

$$(2.10) \quad \mathbb{P}_{x_0}[\tau_{\text{cov}} > 8t_{\text{cov}}(G)] + \max_{v \in G} \mathbb{P}_v[\tau_G > 10t_{\text{mix}}(G)].$$

The first term on the right hand side is at most $1/8$ by Markov's inequality, and for the second term, since τ_G is separation-optimal, (i.e. it is equality in (1.16)), we can put

$$\mathbb{P}_v[\tau_G > 10t_{\text{mix}}(G)] = s_G(10t_{\text{mix}}(G)) \stackrel{\star}{\leq} 4d_G(5t_{\text{mix}}(G)) \stackrel{\triangle}{\leq} 4\left(\frac{2}{4}\right)^5 = \frac{1}{8},$$

uniformly over the starting state v . In the inequality with \star we used Lemma 1.10, and the one with \triangle we used the sub-multiplicativity (1.11). Combining this estimate with (2.10) and (2.9) and the fact that $t_{\text{mix}}(G) \leq t_{\text{hit}}(G) \leq t_{\text{cov}}(G)$ for all reversible chains (see [LPW08, Chapter 10.5, 11.2]), yields that

$$t_{\text{mix}}(\mathbf{Z}_2 \wr G) \leq 8t_{\text{cov}}(G) + 10t_{\text{mix}}(G) \leq 18t_{\text{cov}}(G).$$

This finishes the proof of the upper bound with $C_2 = 18$. \square

Now we turn to investigate the relaxation time of $\mathbf{Z}_2 \wr G$. To do so, we will use Lemma 1.13 and investigate the behavior of $s(t)^{1/t}$ as $t \rightarrow \infty$.

PROOF OF THE UPPER BOUND FOR RELAXATION TIME OF $\mathbf{Z}_2 \wr G$. To prove the upper bound, we will estimate the tail behavior of the strong stationary time $\tau^\diamond = \tau_{\text{cov}}(G) + \tau_G(X_{\tau_{\text{cov}}})$ in Lemma 2.4, relate it to $s^\diamond(t)$, the separation distance on $\mathbf{Z}_2 \wr G$. We will use \mathbb{P} for $\mathbb{P}_{(\underline{f}, x)}$ for notational convenience. Combining (1.16) by union bound we have

$$(2.11) \quad s_{(\underline{f}, x)}^\diamond(t) \leq \mathbb{P}_{(\underline{f}, x)}[\tau^\diamond > t]$$

$$(2.12) \quad \leq \mathbb{P}[\tau_{\text{cov}}(G) > t/2]$$

$$(2.13) \quad + \max_{y \in G} \mathbb{P}_y[\tau_G > t/2]$$

We write τ_w for the hitting time of $w \in G$. We claim that the first term (2.12) can be bounded from above by:

$$(2.14) \quad \mathbb{P}[\tau_{\text{cov}}(G) > t/2] \leq \mathbb{P}[\exists w : \tau_w > t/2] \leq |G|2e^{-\frac{\log 2}{4} \frac{t}{t_{\text{hit}}(G)}},$$

where $t_{\text{hit}}(G)$ is the maximal hitting time of the graph G , see (2.2). To see this, use Markov's inequality on τ_w to obtain that for all starting states $v \in G$ we have $\mathbb{P}_v[\tau_w > 2t_{\text{hit}}(G)] \leq 1/2$, and then run the chain on G in blocks of $2t_{\text{hit}}(G)$. In each block we hit w with probability at least $1/2$, so we have

$$\mathbb{P}_v[\tau_w > K(2t_{\text{hit}}(G))] \leq \frac{1}{2^K}.$$

To get a similar bound for arbitrary t , we can move from $\lfloor t/2t_{\text{hit}}(G) \rfloor$ to $t/2t_{\text{hit}}(G)$ by adding an extra factor of 2, and (2.14) immediately follows by a union bound.

For the second term (2.13) we prove the following upper bound:

$$(2.15) \quad \mathbb{P}_v [\tau_G \geq t/2] \leq |G| e^{-\frac{t}{2t_{\text{rel}}(G)}}.$$

First note that according to Lemma 1.13, the tail of the strong stationary time τ_G is driven by λ_G^t with λ_G being the second largest eigenvalue of the lazy random walk on G . More precisely, using the first line in (1.20) we have that for any initial state $v \in G$:

$$\mathbb{P}_v [\tau_G \geq t/2] \leq s_G(t/2) \leq \frac{1}{\pi_{\min}(G)} \lambda_G^{t/2} \leq |G| \exp \left\{ -\frac{(1 - \lambda_G)t}{2} \right\},$$

where we used that regularity of G implies $\pi_{\min}(G) = |G|^{-1}$, and the inequality $1 - x \leq e^{-x}$ for $x = 1 - \lambda_G$. Next we combine the bounds in (2.14) and (2.15) on (2.11) with the second inequality in (1.20) to estimate the second largest eigenvalue on $\mathbf{Z}_2 \wr G$ as follows:

$$(2.16) \quad |\lambda_2|^t \leq 2d^\circ(t) \leq 2s^\circ(t) \leq 4|G| \exp \left\{ -\frac{\log 2}{4} \frac{t}{t_{\text{hit}}(G)} \right\} + 2|G| \exp \left\{ -\frac{t}{2t_{\text{rel}}(G)} \right\}.$$

In the final step we apply Lemma 1.13: we take the power $1/t$ and limit as t tends to infinity with fixed graph size $|G|$ on the right hand side of (2.16) to get an upper bound on λ_2 . Then we use that $(1 - e^{-x}) \leq x + o(x)$ for small x and obtain the bound on $t_{\text{rel}}(\mathbf{Z}_2 \wr G)$ finally:

$$t_{\text{rel}}(\mathbf{Z}_2 \wr G) \leq \max \left\{ \frac{4}{\log 2} t_{\text{hit}}(G), 2t_{\text{rel}}(G) \right\}.$$

Then, taking into account that $t_{\text{rel}}(G) \leq ct_{\text{mix}}(G) \leq Ct_{\text{hit}}(G)$ holds for any lazy reversible chain (see e.g. [LPW08, Chapter 11.5, 12.2]), we can ignore the second term. \square

PROOF OF THE LOWER BOUND FOR THE RELAXATION TIME. We do not include the proof of the lower bound of the relaxation time in these lecture notes since it is based on a somewhat different technique: it relies on the analysis of the Dirichlet form of the lamplighter walk, with an appropriately chosen test-function f . For more details see [LPW08, Chapter 19.2] for $0-1$ lamps or [KP12] for general lamp graphs. \square

2.2. Generalized lamplighter walks. One can think of a generalisation of lamplighter walks of the following form: instead of $0-1$ lamps, put at each site of the base graph G an identical copy of machine, whose states are represented by a *lamp graph* H with a fixed Markov chain transition matrix Q on H . The walker then does the following: as he follows a simple random walk on the base graph, he modifies the state of the machines along his path randomly according to the transition matrix Q . The state space in this case is a vector of the states of each machine plus the position of the walker. We denote the corresponding graph by $H \wr G$. One step of the lamplighter walk is then: refresh the machine of the departure site, move one step on the base graph, refresh the machine on the arrival site. With this dynamics, one can show that the product measure of the stationary measure of Q over $v \in G$ multiplied by π_G is stationary for this dynamics and the chain is reversible. We denote the resulting graph by $H \wr G$. We can characterise the relaxation time of such walks as follows, from [KP12]:

THEOREM 2.5. *Let us assume that G and H are connected graphs with G regular and the Markov chain on H is lazy, ergodic and reversible. Then there exist universal constants $0 < c_1, C_1 < \infty$ such that the relaxation time of the generalized lamplighter walk on $H \wr G$ satisfies*

$$(2.17) \quad c_1 \leq \frac{t_{\text{rel}}(H \wr G)}{t_{\text{hit}}(G) + |G|t_{\text{rel}}(H)} \leq C_1,$$

THEOREM 2.6. *Assume that the conditions of Theorem 2.5 hold. Then there exist universal constants $0 < c_2, C_2 < \infty$ such that the mixing time of the generalized lamplighter walk on $H \wr G$ satisfies*

$$(2.18) \quad \begin{aligned} c_2(t_{\text{cov}}(G) + t_{\text{rel}}(H)|G|\log|G| + |G|t_{\text{mix}}(H)) &\leq t_{\text{mix}}(H \wr G), \\ t_{\text{mix}}(H \wr G) &\leq C_2 \left(t_{\text{cov}}(G) + |G|t_{\text{mix}}(H, \frac{1}{|G|}) \right). \end{aligned}$$

If further the Markov chain is such that

(A): *There is a strong stationary time τ_H for the Markov chain on H which possesses a halting state $h(x)$ for every initial starting point $x \in H$,*

then the upper bound of (2.18) is sharp.

The proofs above for 0–1 lamps can be modified to work for general lampprographs H . In this case, we also have to construct an ‘almost’ stationary time similar to τ_{cov} and a true stationary time τ^\diamond . The first can be done by using copies of a separation-optimal $\tau_H(v)$, $v \in G$, such that each $\tau_H(v)$ is measured only using the transition steps of the chain on the machine H_v at $v \in G$. Then we wait until all of the $\tau_H(v)$ -s have happened. One can then show that this time is ‘almost’ stationary in the sense that reaching it, the state of the lamp-graphs are stationary, but the position of the walker is not. A similar estimate to that in Lemma 2.3 gives a lower bound on the separation distance. Adding an extra τ_G again gives a ‘true’ strong stationary time τ^\diamond .

In most estimates for the mixing and relaxation time of $H \wr G$ we can use these two stopping times, but there are new terms arising: one has to estimate the local-time structure of the base graph and also the behaviour of τ_H -s. The proofs are worked out in [KP12].

We mention that the upper and lower bound on the mixing time for $H \wr G$ do match for a wide selection of H and G , but not in general. It remains an open problem to give a general formula for the mixing time.

3. Varopoulos-Carne long range estimate

In this section we move on to give a general bound on transition probabilities of SRW on graphs. Later, we will use this estimate to determine the speed of RW on different groups. Let $P = (p(x, y))$ be a transition probability matrix on state space \mathcal{S} . Assume reversibility, i.e., that $\pi(x) > 0$ and $\pi(x)p(x, y) = \pi(y)p(y, x)$ for all $x, y \in \mathcal{S}$.

We may consider \mathcal{S} as the vertex set of an undirected graph where x, y are adjacent iff $p(x, y) > 0$. Let $\rho(x, y)$ denote the graph distance in \mathcal{S} . We assume \mathcal{S} is locally finite (each vertex has finite degree). We now state the *Varopoulos-Carne long-range estimate*:

THEOREM 3.1 (Varopoulos-Carne). $\forall x, y \in \mathcal{S}$ and $\forall t \in \mathbf{N}$,

$$(3.1) \quad p^t(x, y) \leq 2\sqrt{\frac{\pi(y)}{\pi(x)}} \cdot \mathbb{P}(S_t \geq \rho(x, y)) \leq 2\sqrt{\frac{\pi(y)}{\pi(x)}} e^{-\frac{\rho^2(x, y)}{2t}},$$

where (S_t) is simple random walk on \mathbf{Z} .

REMARK 3.2. The Varopoulos-Carne estimate gives good bounds on transition probabilities between vertices that are far away from each other. Another, short-distance estimate is the following, that can be found in various forms in the literature, see e.g. [LPW08, Theorem 17.17]. Let P be the transition matrix of lazy random walk on a graph of maximal degree Δ . Then

$$|P^t(x, x) - \pi(x)| \leq \frac{\sqrt{2}\Delta^{5/2}}{\sqrt{t}}.$$

PROOF OF THEOREM 3.1. We start by reducing to the finite case. Fix t and x . Denote $\hat{\mathcal{S}} = \{z : \rho(x, z) \leq t\}$.

Now $\forall z, w \in \hat{\mathcal{S}}$, consider the modified transition matrix

$$\hat{p}(z, w) = \begin{cases} p(z, w) & : z \neq w \\ p(z, z) + p(z, \mathcal{S} - \hat{\mathcal{S}}) & : z = w \end{cases}$$

Then \hat{p} is reversible on $\hat{\mathcal{S}}$ with respect to π . Since in t steps, the walk started at x cannot exit $\hat{\mathcal{S}}$, it suffices to prove the inequality for $\hat{\mathcal{S}}$ in place of \mathcal{S} , so we may assume that \mathcal{S} is finite.

Let $\xi = \cos \theta = \frac{e^{i\theta} + e^{-i\theta}}{2}$. Taking the t -th power, we see that the coefficients of the binomial expansion are exactly the transition probabilities of SRW on \mathbf{Z} , which gives

$$\xi^t = \sum_{k=-t}^t \mathbb{P}(S_t = k) e^{ik\theta}.$$

By taking the real part, we get

$$(3.2) \quad \xi^t = \sum_{k=-t}^t \mathbb{P}(S_t = k) \cos k\theta.$$

Now denote $Q_k(\xi) = \cos k\theta$. Observe that $Q_0(\xi) = 1, Q_1(\xi) = \xi$, and the identity

$$\cos(k+1)\theta + \cos(k-1)\theta = 2\cos\theta \cos k\theta$$

yields that $Q_{k+1}(\xi) + Q_{k-1}(\xi) = 2\xi Q_k(\xi)$ for all $k \geq 1$. Thus induction gives that Q_k is polynomial of degree k for all $k \geq 1$; these are the celebrated Chebyshev polynomials. Further, since $Q_k(\xi) = \cos(k\theta)$ for $\xi = \cos \theta \in [-1, 1]$ implies the fact that $|Q_k(\xi)| \leq 1$ for $\xi \in [-1, 1]$.

Using the symmetry of cosine function, we can rewrite (3.2) in the form

$$\xi^t = \sum_{k=-t}^t \mathbb{P}(S_t = k) Q_{|k|}(\xi),$$

which is an identity between polynomials. Applying it to the transition probability matrix P on \mathcal{S} , we infer that

$$(3.3) \quad P^t = \sum_{k=-t}^t \mathbb{P}(S_t = k) Q_{|k|}(P)$$

We know that all eigenvalues of P are in $[-1, 1]$. Furthermore, the eigenvalues of $Q_k(P)$ have the form $Q_k(\lambda)$, where λ is an eigenvalue of P , so they are also in $[-1, 1]$. Hence $\|Q_k(P)v\|_\pi \leq \|v\|_\pi$ for any vector v , where $\|v\|_\pi^2 = \sum_{x \in \mathcal{S}} v(x)^2 \pi(x)$. Using this contraction property we can write

$$Q_k(P)(x, y) = \frac{\langle \delta_x, Q_k(P) \delta_y \rangle_\pi}{\pi(x)} \leq \frac{\|\delta_x\|_\pi \|\delta_y\|_\pi}{\pi(x)} \leq \frac{\sqrt{\pi(x)} \sqrt{\pi(y)}}{\pi(x)} = \sqrt{\frac{\pi(y)}{\pi(x)}}.$$

Note that $P^k(x, y) = 0 \ \forall k < \rho(x, y)$ implies $Q_k(P)(x, y) = 0$ for $k < \rho(x, y)$.

Hence, by (3.3), we have

$$p^t(x, y) = \sum_{|k| \geq \rho(x, y)} \mathbb{P}(S_t = k) Q_{|k|}(P)(x, y) \leq \sum_{|k| \geq \rho(x, y)} \mathbb{P}(S_t = k) \sqrt{\frac{\pi(y)}{\pi(x)}},$$

proving the first inequality in (3.1). The second inequality in (3.1) is an application of the well-known Bernstein-Chernoff bound

$$(3.4) \quad \mathbb{P}(S_t \geq R) \leq e^{-R^2/(2t)}.$$

For the reader's convenience we recall the proof. Suppose that $\mathbb{P}(X = 1) = 1/2 = \mathbb{P}(X = -1)$. Then

$$\mathbb{E}(e^{\lambda X}) = \frac{e^\lambda + e^{-\lambda}}{2} = \sum_{k=0}^{\infty} \frac{\lambda^{2k}}{(2k)!} \leq \sum_{k=0}^{\infty} \frac{\lambda^{2k}}{2^k k!} = e^{\lambda^2/2}.$$

Therefore,

$$\mathbb{E}(e^{\lambda S_t}) = (\mathbb{E}(e^{\lambda X}))^t \leq e^{t\lambda^2/2}.$$

Finally, by Markov's inequality,

$$\mathbb{P}(S_t \geq R) = \mathbb{P}(e^{\lambda S_t} \geq e^{\lambda R}) \leq e^{-\lambda R} \cdot e^{t\lambda^2/2}.$$

Optimizing, we choose $\lambda = R/t$, and (3.4) follows. \square

4. Speed of RW on groups and harmonic functions

In this section we characterize the speed of random walk on groups in terms of bounded harmonic functions. For more on this topic see Chapter 13 in [LP15].

Let G be a (finite or countable) group, with finite generating set S . We assume $S = S^{-1}$, and $d = |S|$. Recall the right-Cayley graph on G is given by $x \sim y \Leftrightarrow y \in xS$, and the corresponding simple random walk (SRW) has

$$(4.1) \quad p_{SRW}(x, y) = \begin{cases} \frac{1}{d}, & \text{for } y \in xS, \\ 0, & \text{otherwise.} \end{cases}$$

We define the lazy random walk (LRW) to avoid periodicity issues:

$$(4.2) \quad p(x, y) = \begin{cases} \frac{1}{2}, & \text{for } y = x \\ \frac{1}{2d}, & \text{for } y \in xS, \\ 0, & \text{otherwise.} \end{cases}$$

That is, the transition matrix $P = (P_{SRW} + I)/2$. We call $e \in G$ the origin, and denote ρ the graph distance in G . We write simply $\rho(e, x) = |x|$.

DEFINITION 4.1. The speed of random walk on G is defined as

$$v(G) := \lim_{n \rightarrow \infty} \frac{\mathbb{E}|X_n|}{n} = \text{a.s.} \lim_{n \rightarrow \infty} \frac{|X_n|}{n}.$$

This definition is valid, since the distance is subadditive by the triangle inequality and the transitivity of G :

$$\rho(e, X_{n+m}) \leq \rho(e, X_n) + \rho(X_n, X_{n+m}) \stackrel{d}{=} \rho(e, X_n) + \rho(e, X_m).$$

Taking expectation yields that the expected distance is submultiplicative, hence the speed exist.

The main goal of here is to characterize *when is the speed positive?* But first some examples:

EXAMPLE 4.2. For every d , $v(\mathbf{Z}^d) = 0$. This is easy to see since $(\mathbb{E}|X_n|)^2 \leq \mathbb{E}(|X_n|^2) = \sum_{i=1}^n \mathbb{E}(|Y_i|^2) = n$ by denoting Y_i the independent unit length increment of the walk at step i .

EXAMPLE 4.3. The speed on the infinite d -ary tree Π_d is $v(\Pi_d) = \frac{d-2}{d}$. In each step of the walk, there are $d-1$ edges increasing the distance from the root by $+1$ and exactly 1 edge decreasing the distance, hence the speed is $d-2d$ for non-lazy RW and $\frac{d-2}{2d}$ for lazy RW.

The third example needs some definitions:

DEFINITION 4.4. A state of the lamplighter group G_d on \mathbf{Z}^d is defined as (S, x) where $S \subset \mathbf{Z}^d$ is a finite subset of vertices and $x \in \mathbf{Z}^d$ is the position of a *marker* or lamplighter. Every state in G_d is connected to $2d+1$ other states in G_d : either the marker moves to a uniformly chosen neighbour of x or it switches the lamp at x : i.e. removes x from S if $x \in S$, and adds x to S if $x \notin S$. The origin in this walk is $(\emptyset, \underline{0})$, i.e. all lamps off, marker at the origin.

The set S describes which ‘lamps’ are on, and the marker can switch lamps only along his path. He either moves on the base graph \mathbf{Z}^d or switches the lamp where he currently is.

EXAMPLE 4.5. The speed of the lamplighter walk on G_1 and G_2 is zero, while $v(G_d) > 0$ for $d \geq 3$.

PROOF. For G_1 we can use the marginal distribution of the marker is just a SRW on \mathbf{Z} , hence its range up to time n is whp less than $c\sqrt{n} \log n$. Thus, any state that the lamplighter can reach in n steps has at most only a connected set of on-lamps of size $c\sqrt{n} \log n$. This has distance at most $K\sqrt{n} \log n$ from the origin, since the marker can just walk along its range, switch off each lamp that is on and return to the origin, taking at most $K\sqrt{n} \log n$ steps for some $K > 0$.

For G_2 , the range of SRW on \mathbf{Z}^2 is whp $n/\log n$, so the same argument can be applied to show that the speed is zero.

For $d > 3$, the range of SRW on \mathbf{Z}^d is linear in n , and with positive probability there are going to be a linear number of lamps on, hence the speed is positive, too. \square

Discussion. We see that it is not the growth rate that characterizes the speed: trees and lamplighter groups both grow exponentially. What does characterize the speed? the answer is given by *bounded harmonic functions*.

4.1. Bounded harmonic functions and tail σ -algebras. We start with a definition:

DEFINITION 4.6. We say that a bounded function $u : G \rightarrow \mathbf{R}$ is harmonic for the simple random walk on G if

$$u(x) = \frac{1}{d} \sum_{y \sim x} u(y),$$

that is we have $u = P_{SRW}u = Pu$.

We define the tail σ -algebra as $\mathcal{T} = \bigcap_n \sigma(X_n, X_{n+1}, \dots)$. \mathcal{T} contains all events which are independent of the trajectory up to any fixed finite time. Tail events can easily generate harmonic functions, we list some examples:

- (1) On Π_d , does the RW end up eventually in a given sub-branch of the tree?

$$u_1(x) := \mathbb{P}_x(\text{RW ends in a given sub-branch of the tree})$$

- (2) On G_d with $d \geq 3$, is the lamp at x eventually on?

$$u_2(x) := \mathbb{P}_x(\text{the lamp at } y \text{ is going to be eventually on})$$

$$u_3(x) := \mathbb{P}(\text{the lamps in the subset } A \text{ are all going to be eventually on})$$

One can easily argue that u_1, u_2, u_3 are non-constant by moving the starting point x further and further away from the points / sets under consideration and using transience properties of the marker.

DEFINITION 4.7. We call $f = f(X_0, X_1, X_2, \dots)$ a *tail-function* if changing finitely many values in the trajectory (X_0, X_1, X_2, \dots) does not change the value of f .

CLAIM 4.8. Every tail function generates a bounded harmonic function by

$$u_f(x) = \mathbb{E}_x(f(X_0, X_1, X_2, \dots))$$

for random walk on groups or for lazy chains.

PROOF. We prove it for lazy chains only. First we start with a total variation bound on binomial random variables¹:

$$\begin{aligned}
 \|\text{Bin}(n, \tfrac{1}{2}) - \text{Bin}(n+1, \tfrac{1}{2})\|_{\text{TV}} &\leq 2^{-n-1} \sum_{k=0}^{\lfloor n/2 \rfloor} \left[2 \binom{n}{k} - \binom{n+1}{k} \right] \\
 (4.3) \qquad &= 2^{-n-1} \sum_{k=0}^{\lfloor n/2 \rfloor} \left[\binom{n}{k} - \binom{n}{k-1} \right] \\
 &= 2^{-n-1} \binom{n}{\lfloor n/2 \rfloor} = \frac{1+o(1)}{\sqrt{2\pi n}}.
 \end{aligned}$$

First fix some $\varepsilon > 0$ and pick n large enough such that $\frac{1+o(1)}{\sqrt{2\pi n}} \|f\|_{\infty} < \varepsilon$. Look at two copies of the lazy walk: (X_0, X_1, X_2, \dots) and $(\tilde{X}_0, \tilde{X}_1, \tilde{X}_2, \dots)$. We can then construct a coupling between these two trajectories by using a non-lazy random walk Y , and set $X_n = Y_{\text{Bin}(n, \frac{1}{2})}$ and $\tilde{X}_{n+1} = Y_{\text{Bin}(n+1, \frac{1}{2})}$. The bound in (4.3) and the coupling characterisation of total variation distance (1.4) tells us that we can couple these two trajectories such that $\mathbb{P}(X_n \neq \tilde{X}_{n+1}) \leq \mathbb{P}(\text{Bin}(n, \frac{1}{2}) \neq \text{Bin}(n+1, \frac{1}{2})) \leq \varepsilon$. Hence, we can write

$$\begin{aligned}
 (Pu_f)(x) - u_f(x) &= \mathbb{E}_x(f(X_1, X_2, \dots)) - \mathbb{E}_x(f(X_0, X_1, X_2, \dots)) \\
 &= \mathbb{E}_x(f(\tilde{X}_1, \tilde{X}_2, \dots)) - \mathbb{E}_x(f(X_0, X_1, X_2, \dots)) \\
 &\leq \|f\|_{\infty} \cdot \mathbb{P}(X_n \neq \tilde{X}_{n+1}) \leq \varepsilon,
 \end{aligned}$$

where in the last step we used that if the two trajectories are coupled by time n , then clearly they only differ in finitely many steps, and f is a tail function, hence it takes the same value on $\{X_n = \tilde{X}_{n+1}\}$. Since ε was arbitrary, we get $Pu_f = u_f$, finishing the proof. \square

The reverse direction is also true:

CLAIM 4.9. Every bounded harmonic function u defines a tail function f_u by

$$f_u(X_0, X_1, X_2, \dots) := \limsup_{n \rightarrow \infty} u(X_n).$$

PROOF. Since u is bounded and harmonic, the function $u(X_n)$ is a bounded martingale. Hence, by the martingale convergence theorem we get that it converges. Further, the definitions of the two claims are giving a correspondence between bounded harmonic functions and tail-functions since $u_{f_u}(x) = \mathbb{E}_x(f_u(X_0, X_1, \dots)) = \mathbb{E}_x(\limsup u(X_n)) = u(x)$ by the martingale stopping theorem. \square

We call a σ -algebra \mathcal{F} trivial if $\forall A \in \mathcal{F}, \mathbb{P}_x(A) \in \{0, 1\}$.

We will need the the following equivalence.

THEOREM 4.10. *For random walk on a group, the tail σ algebra \mathcal{T} is trivial if and only if every bounded harmonic function on G is constant.*

PROOF. Suppose first that \mathcal{T} is trivial. Let u be a bounded harmonic function. Then $\limsup u(X_n)$ is a tail function, so it must be constant a.s. By irreducibility, this constant c does not depend on the starting point. Writing $u(x) = \mathbb{E}_x u(X_n)$ and passing to the limit using the bounded convergence theorem proves that $u(x) = c$ for

¹We set $\binom{n}{-1} := 0$.

all x . This direction is valid for any irreducible Markov chain. The other direction is not hard to verify for lazy irreducible Markov chains: Suppose all bounded harmonic functions are constant, and $A \in \mathcal{T}$. Then it is easy to check that $u(x) = \mathbb{P}_x(A)$ is a harmonic function, so the Lévy zero-one law implies that $\mathbb{P}_x(A) \in \{0, 1\}$ for every x . Without assuming Laziness, but using the group structure instead, one can also show that u is harmonic. This can be proved using entropy or via Derriennic's zero-two law [Der76], see Chapter 13 in [LP15] for details. \square

Entropy. To state the next theorem, we need some basic properties of entropy, which we include here for the reader's convenience.

DEFINITION 4.11. The *entropy* of a random variable X with distribution p_x on state space \mathcal{S} is defined as

$$H(X) := \sum_{x \in \mathcal{S}} p_x \log\left(\frac{1}{p_x}\right).$$

and the *relative entropy* of measure P with respect to another measure Q on the same state space \mathcal{S} is defined as

$$D(P|Q) := \sum_x p_x \log\left(\frac{p_x}{q_x}\right).$$

The relative entropy is always nonnegative since $\log t \leq t - 1$ for $t > 0$, hence

$$-D(P|Q) = \sum_{x \in \mathcal{S}} p_x \log\left(\frac{q_x}{p_x}\right) \leq \sum_{x \in \mathcal{S}} p_x \left(\frac{q_x}{p_x} - 1\right) = 0.$$

Finally, the *conditional entropy* is defined as the entropy of the conditional measure $p(x|y) = \frac{p_{xy}}{p_y}$, i.e.

$$H(X|Y) := \sum_{x,y} p(x|y) \log\left(\frac{1}{\log p(x|y)}\right).$$

We write $H(X, Y)$ for the entropy of the joint distribution of (X, Y) . Then it is not hard to see that

$$H(X|Y) = H(X, Y) - H(Y) \leq H(X),$$

since $H(X) + H(Y) - H(X, Y) = D(p_x \cdot p_y | p_{x,y}) \geq 0$ with equality if and only if X and Y are independent. As a corollary we get that for any three random variables X, Y, Z

$$(4.4) \quad H(X|Y, Z = z) \leq H(X|Z = z) \implies H(X|Y, Z) \leq H(X|Z).$$

It can also be shown that the uniform distribution on set \mathcal{S} (with $|\mathcal{S}| = n$) maximizes the entropy:

$$0 \leq D(P_x, U[\mathcal{S}]) = \sum_{x \in \mathcal{S}} p_x \log(np_x) = \log n - H(X).$$

4.2. The Kaimanovich - Vershik - Varopoulos theorem. The next theorem is by Kaimanovich - Vershik ('83) [KV83] and Varopoulos ('85) [Var85].

THEOREM 4.12. *For random walk on a group G , the followings are equivalent:*

- (1) *the speed $v(G) > 0$,*
- (2) *\exists a bounded non-constant harmonic function u on G ,*
- (3) *the entropy of the walk $h = \lim_{n \rightarrow \infty} \frac{H(X_n)}{n} > 0$.*

PROOF. First we show (2) \leftrightarrow (3). Write the joint entropy in two ways:

$$\begin{aligned} H(X_k, X_n) &= H(X_k) + H(X_n|X_k) = H(X_k) + H(X_{n-k}) \\ H(X_k, X_n) &= H(X_n) + H(X_k|X_n) \end{aligned}$$

Rearranging and taking $k = 1$ yields that

$$(4.5) \quad H(X_1) + H(X_{n-1}) - H(X_n) = H(X_1|X_n) = H(X_1|(X_n, X_{n+1}, \dots)),$$

where the last equality is due to the Markov property. Since conditioning on less information increases the entropy (see (4.4)), $H(X_1|(X_n, X_{n+1}, \dots))$ is an increasing function of n . So, the left hand side in (4.5) is also increasing, so we get that $H_n := H(X_n) - H(X_{n-1})$ is decreasing. Hence, $h_n \rightarrow h$ for some $h \geq 0$. So we get, that $\frac{H(X_n)}{n} \rightarrow h$. Now if $h > 0$, then taking $n \rightarrow \infty$ in (4.5) gives $H(X_1|\mathcal{T}) = H(X_1) - h$, that is, conditioning on \mathcal{T} influences the entropy: hence \mathcal{T} can not be trivial. On the other hand if $h = 0$ then $H(X_k|\mathcal{T}) = H(X_k)$ for all k , hence, the tail \mathcal{T} is *independent* of $\{X_1, X_2, \dots, X_k\}$. Thus, it must be trivial itself.

Next we show (3) \leftrightarrow (1). Apply the Varopoulous-Carne estimate on transitive groups to see that $p_n(x) \leq 2e^{-\frac{|x|^2}{2n}}$, and use this estimate on $-\log p_n(x)$ in the definition of $H(X_n)$ to get

$$H(X_n) = \sum_x p_n(x)(-\log(p_n(x))) \geq \sum_x p_n(x)(-\log 2 + \frac{|x|^2}{2n})$$

Rearranging terms and dividing by n yields

$$\frac{\log 2 + H(X_n)}{n} \geq \frac{\mathbb{E}|X_n|^2}{2n^2} \geq \frac{1}{2} \frac{(\mathbb{E}|X_n|)^2}{n^2} = \frac{1}{2} v(G)^2,$$

where we used Jensen's inequality in the last step. Now clearly $v(G) > 0$ implies $\frac{H(X_n)}{n} > 0$.

On the other hand, we can define the spheres $S_k := \{y \in G : |y| = k\}$ and the measure $Q(x) = \frac{2^{-k-1}}{|S_k|}$ if $x \in S_k$ is a probability measure on G . We calculate the relative entropy

$$0 \leq H(Q|P^n) = \sum_x p_n(x) \log \frac{p_n(x)}{Q(x)} \leq \left(\sum_x p_n(x)(|x| + 1) \log 2d \right) - H(X_n),$$

where we used the bound $-\log Q(x) \leq \log(2d)^{k+1}$ since the degree is d . Now dividing by n yields

$$0 \leq \frac{(\mathbb{E}[|X_n| + 1]) \log 2d}{n} - \frac{H(X_n)}{n},$$

and passing to the limit shows that if $h = \lim_n \frac{H(X_n)}{n} > 0$ then the speed is also positive. This finishes the proof. \square

5. Geometric bounds on mixing times

Let G be a (finite or countable) group, with finite generating set S . We assume $S = S^{-1}$, and $d = |S|$. Recall the right-Cayley graph on G is given by $x \sim y \Leftrightarrow y \in xS$, and consider simple random walk on G as in (4.1). Let ρ denote graph distance in G .

THEOREM 5.1. *For simple random walk on $G = \langle S \rangle$,*

(a) *If $|G| < \infty$, then $\mathbb{E}[\rho(X_0, X_n)^2] \geq \frac{n}{2d}$ for $n \leq \frac{1}{1-\lambda}$, where $\lambda = \lambda_2$ is the second eigenvalue.*

(b) *If $|G| = \infty$ and G is amenable, then $\mathbb{E}[\rho(X_0, X_n)^2] \geq \frac{n}{d}$ for all $n \geq 1$.*

REMARK 5.2. (1) The theorem is proved in Lee-Peres [LP13] in the more general setting of random walks on transitive graphs.

(2) Part (b) for Cayley graphs was first discovered by Anna Ersler (unpublished) who relied on a harmonic embedding theorem of Mok.

(3) If G is nonamenable, then we know that $\mathbb{E}\rho(X_0, X_n) \geq cn$, so that $\mathbb{E}[\rho(X_0, X_n)^2] \geq c^2 n^2$ for some constant $c > 0$.

Theorem 5.1 for finite, transitive graphs gives a very general upper bound on relaxation and mixing times of finite groups:

COROLLARY 5.3. Write $\text{diam}(G)$ for the diameter of $G = \langle S \rangle$. Then

$$(5.1) \quad \begin{aligned} t_{\text{rel}}(G) &\leq 2d \cdot \text{diam}(G)^2, \\ t_{\text{mix}}(G) &\leq 2d \cdot \text{diam}(G)^2 \cdot \log |G|. \end{aligned}$$

It is an open problem whether $t_{\text{mix}}(G) \leq Cd \cdot \text{diam}(G)^2$ holds for every transitive finite chain.

PROOF OF COROLLARY 5.3. Apply part (a) of Theorem 5.1 with $n = t_{\text{rel}}(G)$:

$$\text{diam}(G)^2 \geq \mathbb{E}[\rho(X_0, X_n)^2] \geq \frac{t_{\text{rel}}(G)}{2d}.$$

For the second inequality, use [LPW08, Theorem 12.3] stating that $t_{\text{mix}}(G) \leq -\log(\pi_{\min})t_{\text{rel}}(G)$. \square

To prove Theorem 5.1, we use the following key lemma from [LP13] (that is valid for transitive graphs as well). We define the *Dirichlet forms* $Q_n(f) := \langle (I - P^n)f, f \rangle$.

LEMMA 5.4. *For the simple random walk on G as in Theorem 5.1 and any $f \in \ell^2(G)$, we have*

$$\mathbb{E}[\rho(X_0, X_n)^2] \geq \frac{1}{d} \frac{Q_n(f)}{Q_1(f)}.$$

PROOF OF THEOREM 5.1 (FINITE CASE) FROM LEMMA 5.4. In the finite case, take f as an eigenfunction such that $Pf = \lambda f$ with $\|f\|_2 = 1$. Then $Q_n(f) = 1 - \lambda^n$. Using the condition $n < \frac{1}{1-\lambda}$ we can write

$$d \cdot \mathbb{E}[\rho(X_0, X_n)^2] \geq \frac{1 - \lambda^n}{1 - \lambda} = \sum_{j=0}^{n-1} \lambda^{-j} \geq \sum_{j=0}^{n-1} \left(1 - \frac{1}{n}\right)^j \geq \sum_{j=0}^{n-1} \left(1 - \frac{j}{n}\right) \geq \frac{n}{2}.$$

\square

The infinite case is harder and will be proved later.

PROOF OF LEMMA 5.4. Given $f \in \ell^2(G)$, construct $F : G \rightarrow \ell^2(G)$ by $F(x) := \{f(gx)\}_{g \in G}$. Compute (with $X_0 = x_0$)

$$\begin{aligned} \mathbb{E}\|F(X_0) - F(X_1)\|_2^2 &= \mathbb{E} \sum_{g \in G} \|f(gX_0) - f(gX_1)\|_2^2 = \sum_x \sum_y |f(x) - f(y)|^2 p(x, y) \\ (5.2) \quad &= \sum_x \sum_y [(f(x))^2 + (f(y))^2 - 2f(x)f(y)]p(x, y) = 2\langle (I - P)f, f \rangle = 2Q_1(f). \end{aligned}$$

Similarly, $\mathbb{E}\|F(X_0) - F(X_n)\|_2^2 = 2Q_n(f)$. Now, (5.2) implies that

$$\frac{1}{d}\|F(x_0) - F(y)\|_2^2 \leq 2Q_1(f)$$

for any x_0, y with $x_0 \sim y$. Thus, F is Lipschitz with $\text{Lip}(F) \leq \sqrt{2dQ_1(f)}$. Therefore, $2Q_n(f) = \mathbb{E}\|F(X_n) - F(X_0)\|_2^2 \leq (\text{Lip}(F))^2 \mathbb{E}[\rho(X_0, X_n)^2] \leq 2dQ_1(f) \mathbb{E}[\rho(X_0, X_n)^2]$.

Rearranging proves Lemma 5.4. \square

Now we turn to the proof of Theorem 5.1 for infinite G . We will need the following lemma:

LEMMA 5.5. Given $f \in \ell^2(G)$,

$$\mathbb{E}[\rho(X_0, X_n)^2] \geq \frac{n}{d} - \frac{n^2}{2d} \frac{\|(I - P)f\|^2}{Q_1(f)}.$$

PROOF. We use Lemma 5.4. We need to lower bound $Q_n(f)$, and show that it grows almost linearly. For this, we use the differences and bound second differences as follows:

$$\Delta_j = Q_{j+1}(f) - Q_j(f) = \langle P^j f - P^{j+1} f, f \rangle = \langle (I - P)P^j f, f \rangle = \langle P^j f, (I - P)f \rangle.$$

Thus,

$$\begin{aligned} |\Delta_j - \Delta_{j-1}| &= |\langle P^{j-1}(I - P)f, (I - P)f \rangle| \\ &\leq \|P^{j-1}(I - P)f\|_2 \cdot \|(I - P)f\|_2 \leq \|(I - P)f\|_2^2 := \delta, \end{aligned}$$

by Cauchy-Schwarz. Now $\Delta_0 = Q_1(f)$ and $\Delta_j \geq \Delta_0 - j\delta$ whence

$$Q_n(f) = \sum_{j=0}^{n-1} \Delta_j \geq n\Delta_0 - \frac{n(n-1)}{2}\delta \geq nQ_1(f) - \frac{n^2\delta}{2}.$$

Thus,

$$\frac{Q_n(f)}{Q_1(f)} \geq n - \frac{n^2\|(I - P)f\|^2}{2Q_1(f)}$$

and the lemma follows from Lemma 5.4. \square

Proving the theorem for G infinite is harder; we first give the proof under an additional assumption.

ASSUMPTION 5.6. Suppose that $\sum_{j=0}^{\infty} (P^j \mathbf{1}_{\{x_0\}})(x) := \text{Green}(x_0, x)$ is in $\ell^2(G)$.

PROOF OF THEOREM 5.1 (INFINITE CASE) ASSUMING ASSUMPTION 5.6. Note that Lemma 5.5 gives the statement of theorem if we can find a sequence of functions f_k for which $\frac{\|(I-P)f_k\|_2^2}{Q_1(f_k)} \rightarrow 0$.

Let $\{A_k\}$ be a sequence of Følner sets, i.e., $\delta_k := \frac{|\partial_E A_k|}{|A_k|} \rightarrow 0$ as $k \rightarrow \infty$. Here $\partial_E A$ denotes the *edge-boundary* of the set A , i.e. the edges between A and A^c . Write $\psi_k = \mathbf{1}_{A_k}$ and $f_k = \sum_{j=0}^{\infty} P^j \psi_k$. Assumption 5.6 implies that $f_k \in \ell^2(G)$. Note that $(I-P)f_k = \psi_k$ and $f_k(x) = \mathbb{E}_x[\sum_{j=0}^{\infty} \mathbf{1}_{\{X_j \in A_k\}}]$. If $\rho(x, A_k^c) \geq r$, then $f_k(x) \geq r$, so combining these yields

$$\begin{aligned} Q_1(f_k) &= \langle (I-P)f_k, f_k \rangle = \sum_{x \in A_k} f_k(x) \geq r |\{x \in A_k : \rho(x, A_k^c) \geq r\}| \\ &\geq r[|A_k| - d|\partial_E A_k|] = r|A_k|(1 - d\delta_k). \end{aligned}$$

Letting $k \rightarrow \infty$ gives $\liminf_{k \rightarrow \infty} \frac{Q_1(f_k)}{|A_k|} \geq r$ whence $\frac{Q_1(f_k)}{|A_k|} \rightarrow \infty$ since r was arbitrarily large. By Lemma 5.5,

$$\mathbf{E}[\rho(X_0, X_n)^2] \geq \frac{n}{d} - \frac{n^2}{2d} \frac{|A_k|}{Q_1(f_k)}.$$

Letting $k \rightarrow \infty$ proves the theorem assuming Assumption 5.6. \square

Removing Assumption 5.6. For the next lemma, we recall that if P is transient or null-recurrent, then we have the pointwise limit,

$$(5.3) \quad P^i f \rightarrow 0 \quad \text{for every } f \in \ell^2(V).$$

LEMMA 5.7. *Suppose that P satisfies (5.3) and, for some $\theta \in (0, \frac{1}{2})$, there exists an $f \in \ell^2(V)$ with $\|f\|_2 = 1$ and $\|Pf - f\|_2 \leq \theta$. Then there exists a $\varphi \in \ell^2(V)$ such that*

$$(5.4) \quad \frac{\|(I-P)\varphi\|_2^2}{\langle \varphi, (I-P)\varphi \rangle} \leq 32\theta.$$

PROOF OF THEOREM 5.1 FOR INFINITE G WITHOUT ASSUMPTION 5.6. The proof follows by picking $f := \tilde{\psi}_k = \mathbf{1}_{A_k}/\sqrt{|A_k|}$ for the Følner sets defined above. By picking k large enough $\tilde{\psi}_k$ satisfies the condition of Lemma 5.7 for arbitrarily small $\theta > 0$, since in this case $\|f\|_2^2 = |A_k|/|A_k| = 1$ and

$$\|P\tilde{\psi}_k - \tilde{\psi}_k\|_2^2 = \sum_{x \in A_k} \mathbb{E}(X_1 \in A^c) \leq \frac{|\partial_E A_k|}{|A_k|} = \delta_k \rightarrow 0.$$

Combining then these with Lemma 5.5 yields the proof. \square

PROOF OF LEMMA 5.7. Given $f \in \ell^2(V)$ and $k \in \mathbb{N}$, we define $\varphi_k \in \ell^2(V)$ by

$$\varphi_k = \sum_{i=0}^{k-1} P^i f.$$

First, using $(I-P)\varphi_k = (I-P^k)f$ and the fact that P is a contraction, we have

$$(5.5) \quad \|(I-P)\varphi_k\|_2^2 \leq 4\|f\|_2^2.$$

On the other hand,

$$\begin{aligned}\langle \varphi_k, (I - P)\varphi_k \rangle &= \langle \varphi_k, (I - P^k)f \rangle \\ &= \left\langle (I - P^k) \sum_{i=0}^{k-1} P^i f, f \right\rangle \\ &= \langle 2\varphi_k - \varphi_{2k}, f \rangle,\end{aligned}$$

where in the second line we have used the fact that $I - P^k$ is self-adjoint. Combining this with (5.5) yields

$$(5.6) \quad \frac{\|(I - P)\varphi_k\|_2^2}{\langle \varphi_k, (I - P)\varphi_k \rangle} \leq \frac{4\|f\|_2^2}{\langle 2\varphi_k - \varphi_{2k}, f \rangle}.$$

The following claim will conclude the proof.

Claim: There exists a $k \in \mathbb{N}$ such that

$$(5.7) \quad \langle 2\varphi_k - \varphi_{2k}, f \rangle \geq \frac{1}{8\theta}.$$

It remains to prove the claim. By assumption, f satisfies $\|f\|_2 = 1$, and $\|Pf - f\|_2 \leq \theta$. Since P is a contraction, we have $\|P^j f - P^{j-1} f\|_2 \leq \theta$ for every $j \geq 1$, and thus by the triangle inequality, $\|P^j f - f\|_2 \leq j\theta$ for every $j \geq 1$. It follows by Cauchy-Schwarz that $\langle f, (I - P^j)f \rangle \leq j\theta$, therefore

$$\langle f, P^j f \rangle \geq 1 - j\theta.$$

Thus for every $j \geq 1$,

$$\langle \varphi_{2^j}, f \rangle \geq 2^j(1 - 2^j\theta).$$

Fix $\ell \in \mathbb{N}$ so that $2^\ell\theta \leq \frac{1}{2} \leq 2^{\ell+1}\theta$, yielding

$$(5.8) \quad \langle \varphi_{2^\ell}, f \rangle \geq \frac{1}{8\theta}.$$

Now, let $a_m = \langle \varphi_{2^m}, f \rangle$, and write, for some $N \geq 1$,

$$a_\ell - \frac{a_N}{2^{N-\ell}} = \sum_{m=\ell}^{N-1} \frac{2a_m - a_{m+1}}{2^{m-\ell+1}}.$$

By (5.3), we have $\langle P^i f, f \rangle \rightarrow 0$ as $i \rightarrow \infty$, hence $\lim_{N \rightarrow \infty} \frac{a_N}{2^N} = 0$. Using (5.8) and taking $N \rightarrow \infty$ on both sides above yields

$$\frac{1}{8\theta} \leq a_\ell = \sum_{m=\ell}^{\infty} \frac{2a_m - a_{m+1}}{2^{m-\ell+1}}.$$

Since $\sum_{m=\ell}^{\infty} \frac{1}{2^{m-\ell+1}} = 1$, there must exist some $m \geq \ell$ with $2a_m - a_{m+1} \geq \frac{1}{8\theta}$. This establishes the claim (5.7) for $k = 2^m$ and, in view of (5.6), completes the proof of the lemma. \square

6. Balanced random walks with interaction

First we start with some examples.

6.1. Some examples.

EXAMPLE 6.1. A martingale $(X_n)_n$ in \mathbf{Z}^2 , moves horizontally at times $t \in [2^{2k}, 2^{2k+1})$ with k even and vertically $t \in [2^{2k+1}, 2^{2k+2})$ (to nearest neighbours, with equal probabilities in both cases).

Informally, this process is between one and two dimensional, as it has long one-dimensional segments.

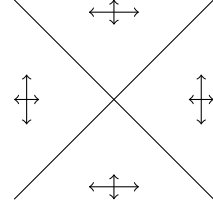
CLAIM 6.2. This process is transient.

PROOF. In the k th horizontal segment, the process can only visit x if it is on the right horizontal line, which has probability $O(1/\sqrt{2^k})$. Since this is summable, the process only visits x finitely many times. Similarly for vertical segments. \square

EXAMPLE 6.3 (Benjamini–Kozma–Schapira [BKS11]). A martingale $(X_n)_n$ in \mathbf{Z}^2 , moves vertically on the first visit to each site, and horizontally on subsequent visits.

QUESTION 6.4. Is this recurrent or transient? [BKS11] includes this and several other open problems of similar nature.

EXAMPLE 6.5 (Nina Gantert; see Ofer Zeitouni’s St. Flour lecture notes on RWRE). On \mathbf{Z}^2 again, a martingale moves horizontally with probability $2/3$ (long arrows) and vertically with probability $1/3$ when $|x| < |y|$, and with opposite probabilities otherwise (including $|x| = |y|$).



PROPOSITION 6.6. This process is transient.

For the proof we use the following basic results.

LEMMA 6.7. If a Markov chain on S has non-constant $\phi : S \rightarrow \mathbf{R}^+$ with $P\phi \leq \phi$ (pointwise) then the chain is transient.

PROOF. $\phi(X_t)$ is a non-negative super-martingale, and so must converge, which contradicts recurrence. \square

LEMMA 6.8 (Excessive measure). If $\mu P \leq \mu$ pointwise and $\mu P \neq \mu$ for a positive measure μ on S , then (X_t) is transient.

PROOF. For any recurrent irreducible chain we have a stationary measure given by $\pi(x) = \mathbf{E}_a \sum_{i=0}^{\tau_a^+ - 1} 1_{X_i=x}$, where τ_a^+ is the return time, and a is an arbitrary reference state. Consider the reverse chain with transitions $\hat{p}(x, y) = \frac{\pi(y)p(y, x)}{\pi(x)}$. Then π is also stationary for \hat{P} . Moreover $P_{xx}^n = \hat{P}_{xx}^n$, and so \hat{P} is also recurrent.

In our case, the assumptions imply that $\phi = \frac{\mu}{\pi}$ has $\hat{P}\phi \leq \phi$. By Lemma 6.7 \hat{P} is transient, and so P must be transient as well. \square

PROOF OF PROPOSITION 6.6. Consider $\mu \equiv 1$. Then $\mu P \leq \mu$ and is strictly smaller at 0. \square

6.2. Walks with few step distributions. [BKS11] raise the following questions.

QUESTION 6.9. Fix two measures μ_1, μ_2 on \mathbf{Z}^d , $d \geq 3$ with mean 0 and bounded support of full dimension. Consider a process that makes steps with law μ_2 on the first visit to a site, and μ_1 on all subsequent visits. When is this recurrent/transient?

QUESTION 6.10. More generally, what if the process moves from X_t by μ_1 or μ_2 and the choice is adapted to \mathcal{F}_t .

The next theorem answers these questions (from [PPS13])

THEOREM 6.11. *Fix any two measures μ_1, μ_2 on \mathbf{Z}^d , $d \geq 3$ with mean 0 and bounded support of full dimension. Let $(X_t)_t$ be a process such that conditioned on \mathcal{F}_t the step $X_{t+1} - X_t$ has law either μ_1 or μ_2 . Then (X) is transient.*

In contrast, there are recurrent processes with three possible step distributions:

EXAMPLE 6.12. In \mathbf{Z}^3 , make a step of ± 1 in the coordinate with maximal absolute value with probability $1 - 2\varepsilon$, and in each of the other coordinates with probability ε each.

THEOREM 6.13. *This process is recurrent for $\varepsilon > 0$ small enough. In \mathbf{Z}^d a similar construction works with d measures.*

Compare this to a continuous diffusion with larger variance in the radial direction. The absolute value is a Bessel process, and by adjusting the covariance matrix, we can control the dimension and even make it less than 2, making the process recurrent. The proof is based on careful construction of a Lyapunov function.

PROOF OF THEOREM 6.11. First we investigate the case of a single increment measure μ . Let Z have law μ , and consider $M = \mathbf{Cov}(\mu) = \mathbb{E}(ZZ^T)$. By applying a linear map, we may assume this is a diagonal matrix $\text{diag}(\lambda)$.

Let $\phi(x) = |x|^{-2\alpha}$. Using a Taylor expansion we have

$$\frac{\phi(x+z)}{\phi(x)} = 1 - \frac{2\alpha x^T z}{|x|^2} - \frac{\alpha|z|^2}{|x|^2} + \frac{\alpha(\alpha+1)}{2} \frac{4x^T z z^T x}{|x|^4} + O(|x|^{-3}).$$

Taking expectation (with $\mathbb{E}(Z) = 0$) we get

$$\begin{aligned} \mathbb{E} \left(\frac{\phi(x+Z)}{\phi(x)} \right) &= 1 + \frac{\alpha}{|x|^4} (-\mathbb{E}|Z|^2|x|^2 + 2(\alpha+1)x^T M x) + O(|x|^{-3}) \\ &= 1 + \frac{\alpha}{|x|^4} \sum_i |x_i|^2 (2(\alpha+1)\lambda_i - \text{tr } M) + O(|x|^{-3}) \end{aligned}$$

If

$$(6.1) \quad 2\lambda_{\max} < \text{tr } M$$

and $\alpha > 0$ is sufficiently small then we get transience, since the sum is negative and dominates the error term. We can truncate ϕ so that the inequality holds for small x as well. Hence, transience follows from Lemma 6.7.

Clearly (6.1) is impossible for 2-dimensional matrices, so we need dimension at least 3.

Note that if there are several increment laws μ_i , the same ϕ may be superharmonic for all of them simultaneously. In that case, an arbitrary adapted choice of μ_i for the steps does not affect transience.

For steps with a single law, we may consider instead the process $M^{-1/2}X$ which has $\mathbf{Cov} = I$, and (6.1) holds.

For a pair of matrices, we can always ensure (6.1), hence transience is guaranteed:

CLAIM 6.14. For any pair of 3×3 symmetric positive definite matrices M_1, M_2 there is an A so that AM_iA^T both satisfy (6.1).

To see this, first apply some A to make M_1 the identity, next diagonalize M_2 by a unitary matrix, (thus keeping $M_1 = I$). If at this point $M_2 = \text{diag}(a, b, c)$ apply $A = \text{diag}(\sqrt{b/a}, 1, 1)$ to finish, as the matrices are now $\text{diag}(b/a, 1, 1)$ and $\text{diag}(b, b, c)$. \square

Acknowledgement. We are grateful to Omer Angel, Jian Ding and Miki Racz for scribing some of these notes, and to Lucas Boczkowski and Perla Sousi for helpful corrections.

References

- [AD86] David Aldous and Persi Diaconis, *Shuffling cards and stopping times*, Amer. Math. Monthly **93** (1986), 243–297.
- [AF02] David Aldous and James Allen Fill, *Reversible markov chains and random walks on graphs*, University of California, Berkeley, 2002.
- [BKS11] Itai Benjamini, Gady Kozma, and Bruno Schapira, *A balanced excited random walk*, C. R. Math. Acad. Sci. Paris **349** (2011), no. 7-8, 459–462. MR 2788390 (2012d:60133)
- [Der76] Yves Derriennic, *Lois zéro ou deux pour les processus de Markov. Applications aux marches aléatoires*, Annales de l’institut Henri Poincaré (B) Probabilités et Statistiques **12** (1976), no. 2, 111–129 (fre).
- [KP12] Júlia Komjáthy and Yuval Peres, *Mixing and relaxation time for random walk on wreath product graphs*, Electronic Journal of Probability **18** (2012), no. 71, 1–23.
- [KV83] Vadim A. Kaĭmanovich and Anatoly M. Vershik, *Random walks on discrete groups: boundary and entropy*, Ann. Probab. **11** (1983), no. 3, 457–490. MR 704539 (85d:60024)
- [LP13] James R. Lee and Yuval Peres, *Harmonic maps on amenable groups and a diffusive lower bound for random walks*, Ann. Probab. **41** (2013), no. 5, 3392–3419.
- [LP15] Russell Lyons and Yuval Peres, *Probability on trees and networks*, Cambridge University Press, 2015, In preparation. Current version available at <http://mypage.iu.edu/~rdlyons/>.
- [LPW08] David A. Levin, Yuval Peres, and Elisabeth L. Wilmer, *Markov chains and mixing times*, American Mathematical Society, 2008.
- [PPS13] Yuval Peres, Serguei Popov, and Perla Sousi, *On recurrence and transience of self-interacting random walks*, Bulletin of the Brazilian Mathematical Society, New Series **44** (2013), no. 4, 841–867 (English).
- [PR04] Yuval Peres and David Revelle, *Mixing times for random walks on finite lamplighter groups*, Electron. J. Probab. **9** (2004), no. 26, 825–845. MR 2110019 (2005m:60007)
- [Var85] Nicholas Th. Varopoulos, *Long range estimates for Markov chains*, Bull. Sci. Math. (2) **109** (1985), no. 3, 225–252. MR 822826 (87j:60100)

DEPARTMENT OF STOCHASTICS, BUDAPEST UNIVERSITY OF TECHNOLOGY AND ECONOMICS, BUDAPEST, HUNGARY

Current address: Eindhoven University of Technology, Eindhoven, The Netherlands

E-mail address: j.komjathy@tue.nl

MICROSOFT RESEARCH

E-mail address: peres@microsoft.com